DOCUMENT RESUME

ED 284 495 HE 020 574

AUTHOR Murray, Harry G.
TITLE Impact of Student Instructional Ratings on Quality of Teaching in Higher Education.
PUB DATE Apr 87
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-25, 1987).
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Faculty; *College Instruction; Educational Quality; Feedback; Higher Education; *Instructional Improvement; *Instructional Innovation; *Student Evaluation of Teacher Performance; Teacher Attitudes; *Teacher Effectiveness

ABSTRACT

The view that student ratings have contributed to improvement of postsecondary teaching is assessed, based on logical argument and personal observation and research evidence from faculty surveys, field experiments, and longitudinal comparisons. Surveys of faculty attitudes indicate generally positive views on the impact of student ratings on instructional improvement, and provide no support for the claim that student ratings have led to increased use of traditional teacher-centered instruction. Field experiments provide clear evidence that feedback from student ratings produces improvement in perceived teaching effectiveness, particularly if student feedback is supplemented by expert consultation. There is evidence that introducing student ratings in an academic unit can produce significant longitudinal improvement in teaching, particularly if ratings are used in salary, tenure, and promotion decisions. It is argued that, innovation is more common today than it was prior to the advent of student ratings and that highly rated teachers are more likely to use nontraditional methods than are teachers receiving lower ratings. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED284495

Impact of student instructional ratings on quality of teaching in higher education

Harry G. Murray

University of Western Ontario

Abstract

Although student evaluation of teaching has had certain undesirable effects, this paper argues that, on balance, student ratings have had a positive impact on the quality of teaching in colleges and universities. This conclusion is supported by evidence from five sources: (1) logical argument, (2) personal observation, (3) surveys of faculty attitudes, (4) field studies involving experimental manipulation of student feedback, and (5) longitudinal comparisons of quality of teaching within a given academic unit. In opposition to the claim that student ratings have discouraged innovation in teaching and led to entrenchment of traditional methods, it is argued that, if anything, innovation is more common today than it was prior to the advent of student ratings, and furthermore, that highly-rated teachers are more likely to use non-traditional methods than are teachers receiving lower ratings.

Paper presented at annual meetings of American Educational Research Association, Washington, April 1987.

HE 020 574

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Harry G Murray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Impact of student instructional ratings on quality of teaching in higher education

Student ratings have gained widespread acceptance over the past 20 years as a measure of teaching effectiveness in North American colleges and universities. Nearly all postsecondary institutions now have some sort of plan for student evaluation of teaching, with the results of used evaluation as diagnostic feedback to instructors and/or as evidence in decisions on faculty retention, tenure, and promotion. In many institutions, student ratings represent the sole form of documentation on quality of teaching.

Given that student ratings have been with us for at least 20 years and, if anything, appear to be increasing in popularity, it is fair to ask whether the use of these ratings has had a positive or a negative impact on the quality of teaching in higher education. In other words, have student ratings improved teaching in colleges and universities, or have they hindered the improvement of teaching? It must be acknowledged at the outset that the question at issue here, like the chicken-egg enigma and the nature-nurture controversy, is one that is fun to discuss but next to impossible to resolve one way or the other. As one would expect, there is a wide range of opposing views on this issue. Students, for the most part, believe that their teaching evaluations are largely ignored, both by individual teachers and by promotion and tenure committees, and thus have no impact whatsoever on quality of teaching (Murray et al., 1982). Many faculty members, on the other hand, believe that the use of student instructional ratings in personnel decisions causes teachers to inflate grades and weaken instructional content in an attempt to "buy" positive evaluations from students. Tom Wilson, my worthy

opponent in this debate, contends that student rating forms imply a traditional, teacher-centred mode of instruction and thus impede progress toward non-hierarchical, student-centred alternatives. My own view is that, despite certain drawbacks, student ratings have had an overwhelmingly positive impact on the quality of postsecondary teaching. My reasons for believing this are based in part on logical arguments, in part on personal observation, and in part on systematic research evidence.

The logical case for student instructional ratings is that since they incorporate evaluative functions that have been found to improve performance in other contexts, such ratings would be expected to improve teaching similarly. For one thing, student ratings provide informative feedback useful for diagnosing instructional strengths and weaknesses. Second, feedback from students can provide the impetus for professional development activities aimed at improved teaching. Third, use of student ratings in salary, promotion, and tenure decisions gives faculty members a tangible incentive for putting time and effort into improvement of teaching. Finally, use of student ratings in tenure and retention decisions provides a selection mechanism whereby better teachers are more likely to be retained by the institution. There are good reasons, then, for expecting that student ratings <u>should</u> lead to improved teaching, particularly if used for both formative and summative purposes.

Consistent with this expectation, personal observation convinces me that quality of teaching at my own institution, the University of Western Ontario, <u>has</u> improved significantly in recent years, and that this improvement has resulted in part from systematic use of student ratings. The teaching I observe in my faculty colleagues of today is far better, on average, than the teaching I received as an undergraduate student in the same university 25 years ago. Today's teachers take teaching more seriously, put more effort

into teaching, plan their courses more systematically, and make the course content more clear and interesting to students. This impression is not unique to me. In a recent survey of senior faculty members at the University of Western Ontario, only 11% of respondents said that classroom teaching was worse today than when they began their careers 15 or more years ago, whereas 41% said that teaching was better today and 48% said there was no difference (Stalker, 1986). I attribute this positive trend to the fact that campus-wide student evaluation of teaching for salary, promotion and tenure purposes has been mandatory at the University of Western Ontario since 1970. Faculty members take teaching seriously because they know that teaching evaluations make a difference in the institutional reward system. Also, contrary to Tom Wilson's thesis, I see no evidence that mandatory use of student ratings has discouraged faculty members from experimenting with non-traditional, student-centred methods of teaching. As elaborated further below, I would guess that, if anything, instructors are more student-centred today than they were prior to the advent of student ratings, and furthermore, that instructors who receive high ratings from students are more innovative and more student-centred in their teaching than instructors who receive lower ratings.

In addition to logical argument and personal observation, systematic research evidence from three different sources, namely faculty surveys, field experiments, and longitudinal comparisons, provides further support for the view that student ratings have contributed to improvement postsecondary of teaching. These three sources of evidence are reviewed in turn below (cf., Murray, 1984).

Faculty surveys

A search of the research literature on faculty attitudes toward student ratings yielded seven different studies in which faculty members were asked one or both of the following questions: "Do student ratings provide useful

feedback for improvement of teaching?" and "Have student ratings led to improved teaching?" Table 1 summarizes the results of these faculty surveys. In the largest survey to date, carried out by Outcalt (1980) at the nine campuses of the University of California, 67% of 4468 respondents said that student ratings had helped them improve the quality of their teaching, and 78% said they had made changes in their teaching as a result of student ratings. Similar results were obtained by Murray et al (1982) at the University of Western Ontario, where 54% of faculty stated that global student ratings provided useful feedback, 65% favored prose comments from students for this purpose, and 78% said that student ratings of specific teaching behaviors were valuable for feedback purposes. Although results vary somewhat from study to study, the general trend in Table 1 is for faculty respondents to agree that student ratings have had a positive impact on quality of teaching.

In a study not listed in Table 1, Ryan, Anderson & Birchler (1980) asked instructors at the University of Wisconsin-Lacrosse to indicate whether student ratings had caused them to change their frequency of use of various instructional methods and practices. Instructors reported significant increases in a number of practices that would normally be viewed as "good teaching" - for example, explicit definition of objectives, availability for consultation, provision of handouts, and prompt return of exams and papers. Unfortunately, instructors also reported increased use of undesirable teaching practices such as watering down of course content, grade inflation, and decreased exam difficulty. In general, faculty members at Lacrosse felt that student ratings had <u>not</u> improved quality of teaching, although this view does not necessarily follow from their profile of reported behavioral changes. Also of interest is the fact that, contrary to Tom Wilson's position that student ratings cause entrenchment of teacher-centred instructional methods,

faculty members at Lacrosse reported that student ratings had led to decreased use of lecturing and increased use of group discussion (as well as increased response to student questions, and increased relevance of content to student interests).

In summary, surveys of faculty attitudes indicate generally positive views on the impact of student ratings on instructional improvement; and, although evidence is limited, provide no support for the claim that student ratings have led to increased use of traditional teacher-centred instruction.

Field experiments

Further evidence of the beneficial effect of student instructional ratings comes from field research in which student feedback is manipulated experimentally. As illustrated in Figure 1, a typical field experiment of this type involves random assignment of teachers to an experimental group that receives mid-term diagnostic feedback from students and a control group that receives no feedback. The two groups are then compared on global end-of-term student ratings to assess the impact of feedback. In a variation on this basic design, McKeachie et al (1980) compared groups of teachers who, at mid-semester, received either no student feedback, a standard computer printout of student item ratings plus norms, or a computer printout supplemented by individual consultation with an expert teacher who interpreted the printout, provided motivational support, and offered specific suggestions for improvement. The three groups differed significantly in end-of-semester student ratings, with the feedback-plus- consultation group showing the highest ratings, the feedback-only group intermediate, and the control group receiving the lowest ratings. In other words, the results indicated that student feedback led to modest improvement of teaching, whereas student feedback supplemented by expert consultation produced much larger gains in

quality of teaching. Cohen (1980) and Menges & Brinko (1986) reached similar conclusions in meta-analytic reviews of student feedback effectiveness. Cohen found that mid-term feedback produced significant improvement in global student ratings in 10 of 22 experimental comparisons. As shown in Figure 1, the mean increment in end-of-term ratings due to student feedback alone was approximately .10 points (3.70 vs. 3.80 on a 5-point rating scale), which corresponds to 8 percentile points; whereas the mean increment due to student feedback plus expert consultation was approximately .33 raw score points or 24 percentile points. Thus an instructor starting at the 50th percentile in student ratings would be expected to improve to the 74th percentile as a result of mid-term diagnostic feedback plus expert consultation. Gains of this magnitude obviously cannot be dismissed as trivial. Also, Overall & Marsh (1979) reported beneficial effects of student feedback plus consultation on criterion measures other than end-of-term ratings, namely student examination performance and planned course enrollment; and Stevens & Aleamoni (1985) showed that effects of student feedback plus follow-up consultation may persist for as long as ten years.

In summary, field experiments provide clear evidence that feedback from student ratings produces improvement in perceived teaching effectiveness, particularly if student feedback is supplemented by expert consultation.

Longitudinal comparisons

Given the various evaluative functions served by student ratings, including feedback, follow-up training, incentive, and selection, it is reasonable to expect that introduction of a student rating program in a particular academic unit should lead to longitudinal improvement in overall quality of teaching over a period of several years. Unfortunately, few if any studies have provided a proper long-term test of this hypothesis. Figure 2

shows department mean student ratings of teaching for the Department of Psychology, University of Western Ontario, for the academic years 1969-70 to 1984-85 inclusive. The same 10-item student rating form has been used annually in this department since the advent of campus-wide student evaluation in 1969. It would appear that, as indexed by student ratings, there has been steady improvement in departmental teaching effectiveness over the years 1970 to 1985. Similar longitudinal gains in mean instructional ratings within a given academic unit have been reported by Gray & Brandenburg (1985) and Pigott & Rosehart (1983), but only the latter study tracked data from the inception of a new ratings program. These findings are consistent with the hypothesis that use of student ratings leads to improvement of teaching, but other interpretations are obviously possible. It may be, for example, that the longitudinal gains shown in Table 2 are due to teacher age or experience rather than student evaluation per se, or are attributable to some totally extraneous factor such as increased "leniency bias" of student ratings across successive years. And even if improvement in teaching can be unambiguously attributed to student evaluation, it is not clear which aspect or function of evaluation is responsible for this improvement. The rating gains plotted in Figure 2 may have resulted from diagnostic student feedback, follow-up instructional development activity, motivational incentive associated with summative use of student ratings, selective retention of better teachers through hiring and tenure decisions, or some combination of these factors. Although teacher selection provides a plausible interpretation of the present results, it fails to account for similar longitudinal gains found by Gray & Brandenburg (1985) for a sample of instructors that remained fixed across years. A further relevant finding, depicted in Figure 3, is that faculty members at Western Ontario tended to improve steadily in rated teaching

effectiveness from the year of initial appointment to the year in which tenure was granted, but then showed a noticeable decline in teaching, followed by a partial recovery. This finding is not easily explained by leniency, age, experience, or feedback-alone interpretations of teaching improvement. It appears that use of student ratings in salary, tenure, and promotion decisions plays an important moderating role in determining longitudinal gains in quality of instruction. Whereas Cohen (1980) identified expert consultation as a necessary prerequisite for reliable effects of student feedback, the present data suggest that summative use of instructional ratings may play a similar role.

In summary, although results are subject to varying interpretation, there is evidence that introduction of student ratings in an academic unit can produce significant longitudinal improvement in teaching, particularly if ratings are used in salary, tenure, and promotion decisions.

Do student ratings impede innovation?

The evidence reviewed above, including faculty surveys, field experiments, and longitudinal comparisons, supports the view that student evaluation has significantly improved the quality of postsecondary teaching. Although research has typically not addressed the issue of which specific aspects of teaching tend to improve in response to student evaluation, it seems reasonable that improvement would be most likely for those teacher characteristics that are assessed by the typical student rating form - that is, characteristics such as explaining clearly, encouraging student participation, giving constructive feedback, and showing enthusiasm in the classroom. Although few would deny the desirability of improvement of these characteristics, it can be argued (e.g., by Tom Wilson) that the items contained in the typical student rating form reflect an authoritarian,

hierarchical style of teaching, and for this reason the widespread acceptance of student ratings in higher education serves to perpetuate a "restrictive and unjust" pedagogy and to impede the development of innovative student-centred or shared-inquiry methods. It is difficult to find empirical evidence relevant to the claim that student ratings impede instructional innovation. My subjective impression, for what it is worth, is that university teachers tended to use "authoritarian" teaching methods 25 years ago, prior to the advent of student ratings, and they continue to do so today. Then as now, lecturing was by a wide margin the preferred method of instruction. As is the case today, books and articles on college teaching written 25 years ago bemoaned the overuse of lecturing and the resistance of faculty to innovation (e.g., Evans, 1967). It would appear, then, that use of teacher-centred methods and resistance to innovation have been part and parcel of university teaching for many years (perhaps for centuries), and have nothing to do with the recent development and use of student instructional ratings.

As a further, informal test of Tom Wilson's thesis, I compared the requirements and teaching methods of ten University of Western Ontario courses I took as a student in the late 1950's with current requirements and methods of the same courses. In all cases, I used the official course outline as the sole source of information on course content. The results of this comparison are not easy to summarize in quantitative terms. My eyeball impression was that reading assignments and writing requirements are lighter today than 25 years ago, but, contrary to Tom Wilson's position, use of student-centred teaching methods is, if anything, slightly more frequent today than in the past. Whereas courses of 25 years ago were characterized by wall-to-wall lecturing, plus heavy doses of exams and papers, today's courses were more likely to include independent study, class discussion, community field work,

and problem-based learning. In yet another eyeball comparison of course outlines, I found taht use of innovative, student-centred teaching methods was more frequent for psychology instructors receiving high student rating scores than for instructors receiving lower ratings. This difference, like the then vs. now comparison, runs directly counter to Tom Wilson's argument that student ratings discourage innovative teaching.

The Keller or PSI method of instruction provides an interesting case study of an innovation in postsecondary education that showed initial promise but failed to gain widespread acceptance. Can the demise of the PSI method be blamed on the implied orthodoxy of student rating forms? Although this claim has some plausibility, I tend to discount it for two reasons. First, surveys of PSI users and department chairs concerning the abandonment of PSI typically do not identify student rating forms as one of the contributing factors. Lloyd and Lloyd (1986) found that practical problems such as cost, time, and administrative hassles were critical in the demise of PSI courses, although difficulty in achieving merit pay, tenure, and promotion while teaching with PSI was also a factor. Knapper's (1986) survey of chairs of Canadian psychology departments pointed to inflated grade distributions, student feelings of isolation from the instructor, and lack of qualified proctors as problems with PSI teaching. Even Keller himself (1985) does not cite student rating forms as a significant "cause of death" in his recent post mortem on the PSI method. Second, even if we acknowledge that the typical lecture-oriented student rating form is inappropriate for PSI courses, and thus may convey the message that PSI teaching is somehow "unusual" or "improper", it is within our power to develop student rating forms specifically geared to any style of teaching we deem acceptable, including the PSI method, and in so doing avoid the implication that lecturing is the only

proper way to teach. One of the departments in my own university has separate student rating forms for six different types of instruction - including lecture, discussion, laboratory, and clinical supervision. The decline of the PSI method may have resulted from the use of inappropriate student rating forms, or from a clash between PSI precepts and faculty views on what constitutes effective teaching, but it seems unlikely that the use of student ratings per se played any significant role.

Conclusions

1. Evidence from five different sources, namely logical argument, personal observation, faculty surveys, field experiments, and longitudinal comparisons supports the conclusion that student instructional ratings have had a positive impact on quality of teaching in higher education.
2. Although data are limited, available evidence fails to support the view that student ratings perpetuate traditional teacher-centred methods and discourage student-centred innovations.

References

Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. Research in Higher Education, 13, 321-341.

Evans, R. I. (Ed.). (1967). Resistance to innovation in higher education. San Francisco: Jossey-Bass.

Gray, D. M., & Brandenburg, D. C. (1985). Following student ratings over time with a catalog-based system. Research in Higher Education, 22, 155-168.

Gross, R. B., & Small, A. C. (1979). A survey of faculty opinions about student evaluations of instructors. Teaching of Psychology, 6, 216-219.

Keller, F. S. (1985). Lightning strikes twice. Teaching of Psychology, 12, 4-8.

Knapper, C. K. (1986). A survey of instructional practices in Canadian psychology departments. Canadian Psychology, 27(1), 51-62.

Lloyd, M. E., & Lloyd, K. E. (1986). Has lightning struck twice? Use of PSI in college classrooms. Teaching of Psychology, 13, 149-151.

McCready, D. J. (1980). Student evaluations of teaching (Research Paper Series No. 8017). Waterloo, Ontario: Wilfred Laurier University, School of Business and Economics.

McKeachie, W. J., Lin, Y. G., Daugherty, M., Moffett, M. M., Neigler, C., Nork, J., Walz, M., & Baldwin, R. (1980). Using student ratings and consultation to improve instruction. British Journal of Educational Psychology, 50, 168-174.

Menges, R. J. (1980). Student evaluations of instruction and faculty morale. Instructional Evaluation, 5(1), 16-18.

Menges, R. J., & Brinko, K. T. (1986). Effects of student evaluation feedback: A meta-analysis of higher education research. Paper presented at American Educational Research Association meetings, San Francisco.

Murray, H. G. (1984). Impact of formative and summative evaluation of teaching in North American universities. Assessment and Evaluation in Higher Education, 9, 117-132.

Murray, H. G., Newby, W. G., Bowden, B., Crealock, C., Gaily, T. D., Oswin, J., & Smith, P. (1982). Evaluation of teaching at the University of Western Ontario. Provost's Advisory Committee on Teaching and Learning, University of Western Ontario.

Ory, J. C., & Braskamp, L. A. (1981). Faculty perceptions of the quality and usefulness of three types of evaluative information. Research in Higher Education, 15, 271-282.

Outcalt, D. L. (Ed.). (1980). Report of the task force on teaching evaluation. The University of California.

Overall, J. U., & Marsh, H. W. (1979). Mid-term feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. Journal of Educational Psychology, 71, 856-865.

Owens, R. E. (1977). Evaluating the importance of teaching. Manhattan, Kansas: Kansas State University, Office of Educational Research. (ERIC Document No. ED 160 016).

Pigott, A., & Rosehart, R. G. (1983). Development and use of student course evaluations at Lakehead University. Paper presented at conference on Evaluation and Improvement of University Teaching, Montebello, Quebec.

Ryan, J. J., Anderson, J. A., & Birchler, A. B. (1980). Student evaluation: The faculty responds. Research in Higher Education, 12, 317-333.

Stalker, R. G. (1986). Is the quality of university education declining? Survey of faculty attitudes and longitudinal comparison of undergraduate honours theses. Unpublished honours thesis, University of Western Ontario, London, Canada.

Stevens, J. J., & Aleamoni, L. M. (1985). The use of evaluative feedback for instructional improvement: A longitudinal perspective. Instructional Science, 13, 285-304.

Table 1

Surveys of faculty attitudes on formative impact of student instructional ratings

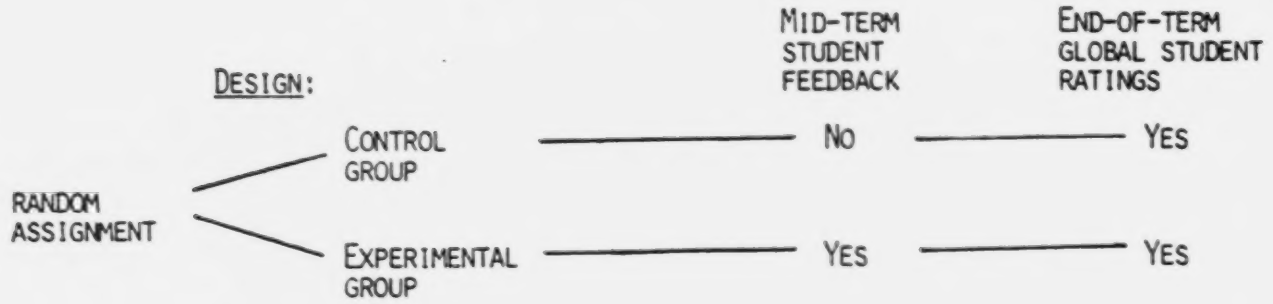
Survey	N	PERCENT AGREEING: Do Student Ratings Provide Useful Feedback?	PERCENT AGREEING: Have Student Ratings Led To Improved Teaching?
McCready (1980) Wilfred Laurier U.	25	76	80
Outcalt (1980) U. California	4468		67
Gross & Small (1979) George Mason U.	163		84
Menges (1980) Northwestern U.		73	
Owens (1977) Kansas State U.			88
Murray et al (1982) U. Western Ontario	666	54 (global ratings) 65 (prose comments) 78 (specific ratings)	
Ory & Braskamp (1981) U. Illinois	25 22	54* (rating scales) 63* (prose comments)	

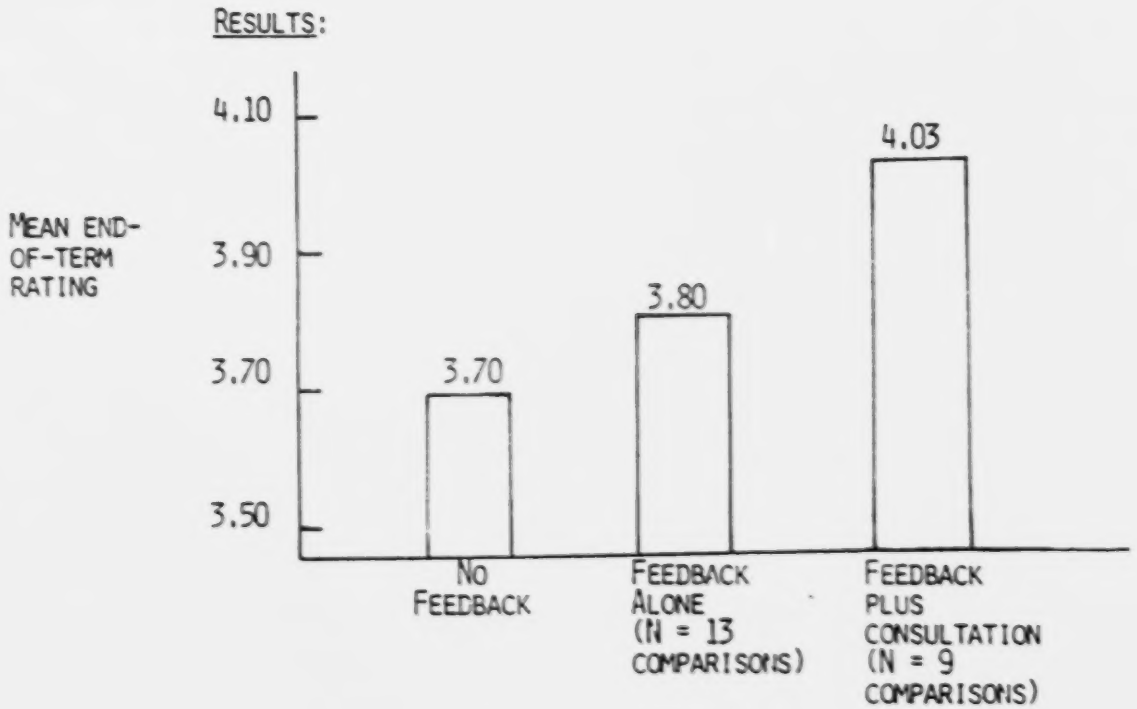
*Estimated from mean ratings on 5-point scale.

Figure 1

Field experiments on effectiveness

of student-rating feedback





SOURCE: COHEN (1980)

Figure 2

Mean student rating of teaching,

Department of Psychology, UWO

for academic years 1969-70 to 1984-85

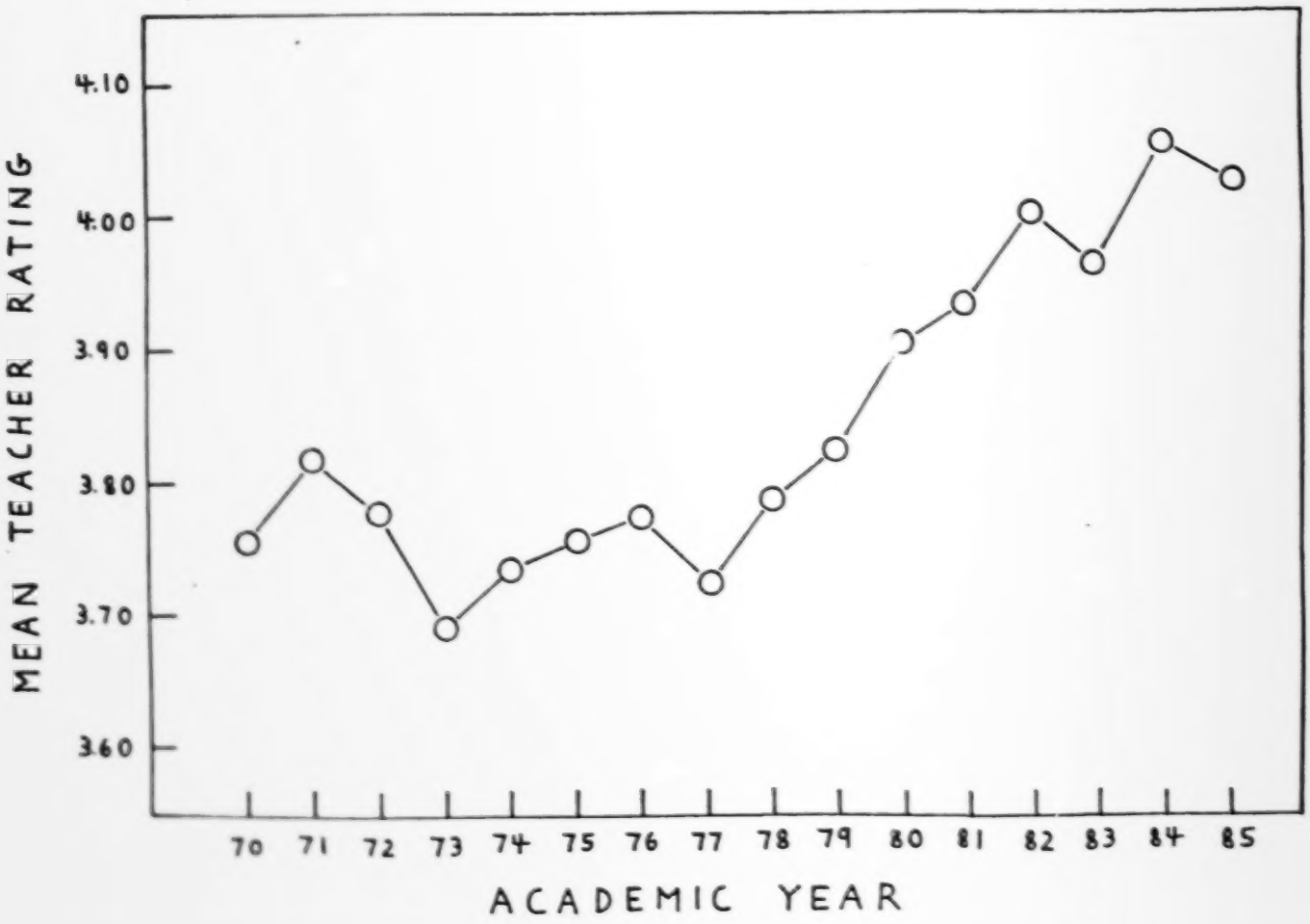


Figure 3

Mean teacher ratings in three pre-tenure and four post-tenure years for faculty members granted tenure between 1972 and 1977 (N=13)